Inter-subject neural code converter for visual image representation.

AUTHOR(S):

Yamada, Kentaro; Miyawaki, Yoichi; Kamitani, Yukiyasu

CITATION:

Yamada, Kentaro ...[et al]. Inter-subject neural code converter for visual image representation.. NeuroImage 2015, 113: 289-297

ISSUE DATE:

2015-04-02

URL:

http://hdl.handle.net/2433/198901

Inter-subject neural code converter for visual image representation

Kentaro Yamada[1,2], Yoichi Miyawaki[3,2,4] , and Yukiyasu Kamitani[2,5]*

[1]Fundamental Technology Research Center, Honda R&D Co.,Ltd., Saitama 351-0188, Japan

[2]ATR Computational Neuroscience Laboratories, Kyoto 619-0288, Japan

[3]National Institute of Information and Communications Technology, Kyoto 619-0288, Japan

[4]The University of Electro-Communications, Tokyo 182-8585, Japan

[5]Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan

*Corresponding author: Yukiyasu Kamitani, Ph.D.

ATR Computational Neuroscience Laboratories

2-2-2 Hikaridai, Keihanna Science City, Kyoto 619-0288, Japan.

Phone: +81-774-95-1212, Fax: +81-774-95-2705.

E-mail: kmtn@atr.jp

Abstract

Brain activity patterns differ from person to person, even for an identical stimulus. In functional brain mapping studies, it is important to align brain activity patterns between subjects for group statistical analyses. While anatomical templates are widely used for inter-subject alignment in functional magnetic resonance imaging (fMRI) studies, they are not sufficient to identify the mapping between voxel-level functional responses representing specific mental contents. Recent work has suggested that statistical learning methods could be used to transform individual brain activity patterns into a common space while preserving representational contents. Here, we propose a flexible method for functional alignment, "neural code converter," which converts one subject's brain activity pattern into another's representing the same content. The neural code converter was designed to learn statistical relationships between fMRI activity patterns of paired subjects obtained while they saw an identical series of stimuli. It predicts the signal intensity of individual voxels of one subject from a pattern of multiple voxels of the other subject. To test this method, we used fMRI activity patterns measured while subjects observed visual images consisting of random and structured patches. We show that fMRI activity patterns for visual images not used for training the converter could be predicted from those of another subject where brain activity was recorded for the same stimuli. This confirms that visual images can be accurately reconstructed from the predicted activity patterns alone. Furthermore, we show that a classifier trained only on predicted fMRI activity patterns could accurately classify measured fMRI activity patterns. These results demonstrate that the neural code converter can translate neural codes between subjects while preserving contents related to visual images. While this method is useful for functional alignment and decoding, it may also provide a basis for brain-to-brain communication using the converted pattern for designing brain stimulation.

Introduction

Human brains are individually unique. Anatomical structure and brain activity patterns differ from person to person, even in response to identical sensory inputs. Anatomical normalization morphs the anatomical structure of one's brain to fit a template brain, for which the Talairach template (Talairach and Tournoux, 1988) and the MNI template (Evans et al., 1992, 1993) are widely used. Anatomical normalization aligns the 3D structure of the brain to a template by an affine transformation, often combined with nonlinear warping. A more precise method for anatomical normalization is to align cortical surfaces between individuals. Since the cortex is a 2D folded surface, cortical surface-based alignment has an advantage over 3D structure-based methods that do not explicitly take the cortical surface information into account. Cortical surface-based alignment utilizes cortical surface geometry such as manually-defined anatomical landmarks (Van Essen, 2004, 2005) and curvature of the cortical surface (Fischl et al., 2008) to calculate the correspondence between individual cortical surfaces.

Although these methods can normalize anatomical structure of brains so that brain activity at corresponding coordinates can be compared between different subjects (Hasson et al., 2004), there still remain differences in the finer spatial patterns of brain activity, presumably originating from idiosyncratic neural representations at a mesoscopic scale. Assimilation of individual differences in brain activity patterns may enable more precise group statistical analyses than normalization of anatomical structure. In addition, it allows us to predict one subject's brain activity patterns from another's. Inter-subject brain activity conversion is potentially useful when constructing a "decoder," a statistical model that predicts a subject's perception and/or behavior from brain activity patterns (Kamitani and Tong, 2005), since decoder training typically needs repetitive measurements of brain activity corresponding to a subject's perception or behavior. Inter-subject brain activity conversion hence may help to reduce the time and cost of performing actual experiments for training decoders for each subject in a study.

Recent studies have proposed methods to assimilate individual differences in brain activity patterns. Sabuncu et al. (2009) developed a method that spatially warps cortical surface points for each subject such that fMRI signals at the same points on the cortical surface are the most correlated. They found that the method could improve group statistics analyzed by a standard general linear model. Haxby et al. (2011) defined a common brain activity space calculated by Procrustean transformation, a combination of orthogonal transformations (rotation, translation, and uniform scaling), of fMRI activity patterns of multiple subjects for an identical sequence of stimuli, and then converted the fMRI activity patterns of each subject into the common brain activity space. They demonstrated that a decoding model trained with the converted fMRI activity patterns was able to achieve high classification performance of object categories.

Here we propose a method to design a "neural code converter," which provides a direct and flexible inter-subject conversion of brain activity patterns. It is a machine learning-based method that aims to predict the brain activity pattern of one person from another's. Our method uses a linear regression model to make a prediction of the intensity of each voxel for a person ("target subject") from multiple voxels of another person ("source subject"), given the same stimulus. This model is capable of representing all kinds of linear transformations and is not limited to orthogonal ones.

In addition, we introduced a Bayesian sparseness constraint on the weights of source voxels (Bishop, 2006) that automatically selects a small number of voxels relevant to the prediction. This approach reduces the dimensionality of the model and is expected to improve prediction accuracy while avoiding overfitting. Further, the efficiency of our neural code converter rests, in part, on biologically plausible sparsity assumptions. In other words, we assume that functional anatomy shows a degree of segregation such that neuronal representations are spatially compact. These sparsity assumptions underlie multivariate Bayesian analyses of distributed responses in fMRI (Yamashita et al., 2008; Friston et al., 2008; Chadwick et al 2014).

To evaluate the conversion performance, we used the fMRI data from our previous study (Miyawaki et al., 2008), in which visual images consisting of 10 x 10 patches ("pixels") were presented. The converter was trained on data from a pair of subjects collected from an identical sequence of presented random images. It was then tested on fMRI data for independent images containing a simple geometric shape. The conversion performance was quantified by the correlation between the measured voxel patterns of the target subject and those predicted by the converter from the source subject. The results are compared with those from other methods using only anatomical information, one-to-one voxel mapping by correlation, retinotopy, or orthogonal transformation of multi-voxel space. We also show that visual images can be reconstructed from brain activity predicted by the neural code converter. Further, we created a decoder for the shapes using converted brain activity to achieve accurate classification of measured data. The results demonstrate the feasibility and usability of the neural code converter for predicting brain activity for unseen stimuli. A preliminary version of this study was presented in conference proceedings (Yamada et al., 2011).

Methods

Algorithms

(1) *Neural code converter*

As shown in Fig. 1, the neural code converter from the source to the target subject consists of multiple linear regression models. Each regression model predicts the amplitude of the target subject's fMRI activity for each voxel. The amplitude of the fMRI activity patterns of the i-th voxel, y_i, is modeled as

$$y_i = \mathbf{w}_i^{\mathrm{T}} \cdot \mathbf{x}, \tag{1}$$

where $\mathbf{x} = [x_1, x_2, ..., x_N]^{\mathrm{T}}$ (N, number of voxels) denotes a set of the amplitude of fMRI activity patterns for all voxels in a region-of-interest (ROI) of the source subject and $\mathbf{w}_i = [w_{i1}, w_{i2}, ..., w_{iN}]^{\mathrm{T}}$ is a weight vector to predict the amplitude of the i-th voxel of the target subject from the multi-voxel patterns of fMRI activity of the source subject. The neural code converter consists of a set of the weight vectors for all voxels in the ROI of the target subject.

The number of fMRI data samples for training is less than the number of voxels in the ROI, or the feature dimension of the model, and thus overfitting is likely to occur. In addition, it is natural to assume that the activity of a particular target voxel would be similar only to a small part of the corresponding areas in the ROI of the source subject because the functional organization of the brain is largely correlated between subjects, especially in the early visual area. Thus only a small number of source voxels may be relevant to prediction for a particular target voxel and the others can be pruned off as irrelevant voxels. To reduce the feature dimension of the model (the number of source voxels), we adopted a variant of sparse regression (Bishop, 2006; Toda et al., 2011), which performs hierarchical Bayesian estimation of the weight parameters with an ARD (automatic relevance determination) prior. The likelihood function of the weight parameters are described as

$$P(\mathbf{y} \mid \mathbf{W}, \mathbf{x}) \propto \exp\left[-\frac{1}{2} \beta \sum_{k=1}^{K} \left\| \mathbf{y}(k) - \mathbf{W}^{\mathrm{T}} \cdot \mathbf{x}(k) \right\|^2 \right], \tag{2}$$

where $\mathbf{y} = [y_1, y_2, \mathrm{L}\ , y_M]^{\mathrm{T}}$ (M ,number of target voxels), $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \mathrm{L}\ , \mathbf{w}_M]$, and β represents an inverse variance of the observation noise. The likelihood function was combined with a prior distribution

for each weight parameter to obtain the posterior distribution. The prior distributions for a weight parameter and a noise variance are described as

$$P(w_{ij} \mid \alpha_{ij}) = \mathrm{N}\left(0, \alpha_{ij}^{-1}\right), \tag{3}$$

and

$$P(\beta) = \beta^{-1}, \tag{4}$$

respectively, where N represents a normal distribution, w_{ij} represents a weight parameter for *i*-th source voxel and *j*-th target voxel, α_{ij} is the hyperparameter denoting the inverse of the variance, or precision, of the weight parameter w_{ij}, and β is the inverse of variance of the observation noise. The hyperparameter α_{ij} was also treated as a random variable, whose distribution is defined by

$$P(\alpha_{ij}) = \alpha_{ij}^{-1}. \tag{5}$$

The weight parameter w_{ij}, the inverse variance of the weight parameter α_{ij}, and inverse variance of the observation noise β were estimated by taking the expectation of the posterior distribution. Since a direct evaluation of the posterior distribution is analytically intractable, we used a variational Bayesian method to approximate the distribution. See Sato (2001), Sato et al. (2004), and Ting et al. (2005, 2008) for details about the algorithm used for parameter estimation. The source codes are available at the websites (http://www.cns.atr.jp/dni/en/downloads/neuralcodeconverter/ and https://github.com/ATR-DNI/NeuralCodeConverter/).

(2) "*One-to-one" conversion*

The neural code converter was designed to receive fMRI signals from multiple voxels of the source subject to predict the fMRI signals of a single voxel for the target subject ("many-to-one" conversion). To examine whether the neural code converter exploits multi-voxel patterns of fMRI activity for the prediction, we also tested the following two types of conversion algorithms, both of which were designed to convert fMRI signals based on the correspondence between each single voxel of the source and target subject ("one-to-one" conversion).

(2-1) *Anatomy-based conversion*

The first method was based on anatomical normalization. This conversion assumed that individual

differences in fMRI activity patterns corresponding to identical stimulus inputs were only caused by differences in the anatomical structure of the brain. If this is the case, individual differences in fMRI activity patterns can be assimilated by matching the anatomical structure of the brains of different subjects. For this purpose, we performed nonlinear anatomical normalization of structural images of the source and target subjects, and applied the same normalization to functional images of both subjects. This procedure converted fMRI activity patterns of the source and target subjects into a common, normalized coordinate space. After this conversion, the fMRI activity of each voxel from the source subject in the common coordinate space was considered as that of the corresponding voxel of the target subject in the common coordinate space.

(2-2) *Correlation-based conversion*

The second method was based on similarity between time-courses of fMRI activity for pairs of voxels between subjects. In this method, we assumed that individual differences in fMRI activity patterns corresponding to identical stimulus inputs were only caused by differences in the relative spatial relationships among voxels within a subject. The conversion was performed by associating a time-course of fMRI activity for each voxel of the target subject with that from the source subject, such that fMRI activity of the paired voxels were most correlated. Then, the intensity of each voxel in the target subject was replaced with that of the associated voxel in the source subject.

(3) *Retinotopy-based conversion*

Although the neural code converter was designed to exploit information represented in the multi-voxel patterns of fMRI signals from the source subject to predict fMRI signals for the target subject, the prediction could be using only the inter-subject correspondence of the well-known functional structure of the early visual area, i.e., retinotopic map only. To examine this possibility, we also tested a retinotopy-based method, in which the intensity of each voxel of the target subject was predicted from the averaged fMRI signal among voxels of the source subject that had the same retinotopy coordinate as the voxel from the target subject.

(4) *Procrustean conversion*

Haxby et al. (2011) showed that Procrustean transformation, a restricted version of linear transformation that involves only translation, rotation, and uniform scaling, was useful for aligning fMRI activity patterns of multiple subjects into a common brain activity space. The neural code converter is a combination of linear weights for source voxel signals to predict target voxel signals, yielding a general linear transformation matrix that converts fMRI activity patterns of the source subject to the target subject, which is a more flexible transformation than Procrustean transformation. To examine whether the flexibility of the neural code converter contributes to the conversion accuracy of the fMRI activity

patterns between subjects, we also tested Procrustean transformation of the fMRI activity patterns of the source subject to the target according to Haxby et al. (2011)'s methods.

fMRI data

To evaluate these algorithms, we used fMRI data from three subjects (one female, two males), one of which was from our previous work (Miyawaki et al., 2008) and the rest were newly collected for this study using a similar protocol to the previous study. All three subjects gave written informed consent and the study was approved by the Ethics Committee of ATR. The fMRI data are available at the websites (http://www.cns.atr.jp/dni/en/downloads/neuralcodeconverter/ and https://github.com/ATR-DNI/NeuralCodeConverter/).

According to our previous work, three types of experimental sessions were performed to measure the fMRI activity patterns of the visual cortex: (1) a random image session, (2) a figure image session, and (3) a retinotopic mapping session (Engel et al., 1994; Sereno et al., 1995).

In the random image session, images of random patterns consisting of 12×12 binary, contrast-defined square patches (1.15×1.15 deg each) were presented with a fixation point at the center of each image on a gray background. Each square patch was either a flickering checkerboard (spatial frequency, 1.74 cycles/deg; temporal frequency, 6 Hz) or a homogeneous gray area, with a probability of 0.5. To avoid effects from the stimulus frame, only the central 10×10-patch area was used for analysis. Each stimulus block consisted of 6-s presentation of a random pattern image followed by a 6-s rest period, and 22 stimulus blocks constituted one run. At the beginning and the end of each run, an extra rest period was inserted for 28 s and 12 s, respectively. In total, twenty runs were performed and a total of 440 different random images were presented for each subject. The same image set was presented in the same order to all subjects. The duration of stimulus presentation was shorter than that used in a typical block design paradigm because .this experiment was designed to obtain fMRI signals for a variety of visual inputs within a limited time to train accurate statistical models.

In the figure image session, figure images in stimulus blocks consisted of flickering checkerboard patches, as in the random image session. The patches in the figure images formed five different types of geometric shapes: a square, small frame, large frame, a plus sign, and the letter X. Each stimulus block lasted for 12 s, followed by a 12-s rest period, and ten stimulus blocks constituted one run. Each geometric shape was presented twice per run. At the beginning and end of each run, an extra rest period of the same duration as the random image session was inserted. To ensure alertness to the stimulus images, each subject was

instructed to detect changes in the color of the fixation point (from red to green, lasting 100 ms).

The retinotopic mapping session was performed using conventional methods, as in Engel (1994) and Sereno et al. (1995). A rotating wedge and an expanding ring consisting of a flickering checkerboard pattern were used as stimulus images. fMRI scanner settings for all sessions were identical to those in Miyawaki et al. (2008) .

MRI data for three experimental sessions explained above were all obtained using a 3.0-Tesla Siemens MAGNETOM Trio A Tim scanner located at the ATR Brain Activity Imaging Center. An interleaved T2*-weighted gradient-echo echo-planar imaging (EPI) scan was performed to acquire functional images to cover the entire occipital lobe (TR, 2000 ms; TE, 30 ms; flip angle, 80deg; FOV, 192×192 mm; voxel size, 3×3×3 mm; slice gap, 0 mm; number of slices, 30). T2-weighted turbo spin echo images were scanned to acquire high-resolution anatomical images of the same slices used for the EPI (TR, 6000 ms; TE, 57 ms; flip angle, 90 deg; FOV, 192×192 mm; voxel size, 0.75×0.75×3.0 mm). T1-weighted magnetization prepared rapid-acquisition gradient-echo (MP-RAGE) fine-structural images of the whole-head were also acquired (TR, 2250 ms; TE, 2.98 or 3.06 ms; TI, 900ms; flip angle, 9 deg; FOV, 256×256mm; voxel size, 1.0×1.0×1.0mm).

Before performing inter-subject conversion, raw fMRI signals were preprocessed as follows. First, signals in each experimental session were processed with standard procedures including motion correction, outlier rejection, detrending, and high path filtering. Then, baseline correction for each voxel was performed using the average signal intensity in the beginning rest period of each session. Each volume was labeled by stimulus after shifting the signal time course by two volumes (4 s) to account for a hemodynamic delay. The signals were simply averaged within each stimulus block to create data samples for machine learning analyses.

Inter-subject conversion procedure

We used only fMRI data for the random image session to train a converter according to each algorithm. After training, each converter was tested with fMRI data for the figure image session, which were independent of the data used for training.

For all algorithms except the anatomy-based conversion, bilateral V1 was used as a ROI to evaluate the conversion accuracy of each algorithm. For anatomy-based conversion, we used WFU PickAtlas (http://fmri.wfubmc.edu/software/PickAtlas) to define Brodmann area 17 (BA17), an area anatomically

equivalent to V1, as a ROI.

fMRI activity patterns from the ROI were preprocessed using the same procedures as in Miyawaki et al. (2008), and were used for each algorithm to convert those from one subject (source subject) to predict those of the other subject (target subject).

The conversion accuracy of each algorithm was evaluated by the spatial correlation between a predicted voxel pattern and a measured voxel pattern of the target subject, which we refer to as the voxel-wise correlation, for each image presented in the figure image session. Brain activity varies across different experimental blocks even for the same stimulus in the same subject. Performance evaluation of inter-subject conversions would be strongly affected by those variances. We hence calculated the correlation between measured brain activity patterns of each target subject for the same stimulus pair presented in different experimental blocks, and used it as an upper bound of the voxel-wise correlation.

Visual image reconstruction

If fMRI activity patterns of a target subject can be predicted accurately, we should also be able to reconstruct visual stimulus images corresponding to the predicted fMRI activity patterns (instead of using measured fMRI activity patterns). Here we evaluated the performance of the neural code converter by reconstructing visual stimulus images from fMRI activity patterns of the target subject predicted from measured fMRI activity patterns of the source subject.

To accomplish this, first, a reconstruction model was trained with fMRI activity patterns of the target subject measured for the random pattern images presented in the random image session. Then, using the neural code converter, fMRI activity patterns of the target subject were predicted from those of the source subject measured for the figure images presented in the figure image session. Finally, visual stimulus images were reconstructed from the predicted fMRI activity patterns for the target subject by using the trained reconstruction model (see Fig. 2).

For the sake of comparison, we also performed conventional visual image reconstruction (Miyawaki et al., 2008) using fMRI data actually measured for the target subject. The original reconstruction method predicted the contrasts of ‘‘local image bases’’ at multiple scales (1×1, 1×2, 2×1, and 2×2 patch areas, defined by rectangles), and the local image bases were multiplied by the predicted contrasts to reconstruct visual images (Miyawaki et al., 2008). Here we predicted the contrasts of only 1×1 local image bases instead of using multiple scales to reconstruct visual images from measured fMRI activity patterns. Note that visual image reconstruction using predicted fMRI activity patterns was also based on the 1×1 scale. We evaluated the performance of the neural code converter by the spatial correlation between images

reconstructed from the predicted and the measured fMRI activity patterns, which we refer to as the pixel-wise correlation. Upper bounds of pixel-wise correlations were also caluculated in a way similar to the voxel-wise correlation.

Decoder trained with converted brain activity

The accuracy of the neural code conversion was further evaluated by performance of a decoding model trained with the predicted fMRI activity patterns for the target subject. We performed a classification analysis where measured fMRI activity patterns of the target subject were classified into a category associated with one of stimulus classes, using a decoder trained with predicted fMRI activity patterns.

To accomplish this, first, the fMRI activity patterns of the target subject for the figure images presented in the figure image session were predicted from those of the source subject for the corresponding stimuli using the neural code converter. Then a classification model was trained with the predicted fMRI activity patterns of the target subject along with the class labels for the corresponding stimuli. Finally, the trained classifier was used to classify fMRI activity patterns of the target subject measured for the corresponding figure images (see Fig. 3). We used an L2-loss support vector machine (LIBLINEAR, http://www.csie.ntu.edu.tw/~cjlin/liblinear/) to create the classification model. The classifier was trained with fMRI activity patterns predicted for three out of four runs in the figure image session and was tested with the remaining run. This evaluation was repeated until all runs were tested (leave-one-run-out cross-validation). For the sake of comparison, we also performed conventional classification analysis (Kamitani and Tong, 2005) using measured fMRI activity patterns of the target subject for both the training and testing. The performance was evaluated based on leave-one-run-out cross-validation in a similar way to the case using the predicted fMRI activity patterns for the training.

Results

First, we evaluated how accurately brain activity patterns were predicted through the neural code converter by comparing them with measured brain activity patterns. We calculated the spatial correlation between predicted and measured voxel patterns for the target subject, which we call the voxel-wise correlation, for each image presented in the figure image session. Since each of the five figure images was presented in eight trials, correlations were calculated for all possible combinations of the trials, yielding 320 correlations (= 5 images×(8×8 trials)) for each source–target subject pair. This analysis was performed for a total of six source–target subject pairs created from the three subjects. Fig. 4 shows the averaged correlation (Fisher's Z-transformed) between predicted and measured brain activity for each source–target pair. The voxel-wise correlation was significantly higher than zero (t-test, $p < 0.05$) in all source–target pairs and the average of the voxel-wise correlations across subjects was 0.399 ± 0.014 (mean ± 95% confidence interval). Compared with upper bound correlations, the neural code converter shows a moderate level of prediction accuracy.

To see how the neural code converter learned the patterns for prediction, we analyzed the weight values of the voxels of the source subject (source voxels) for each voxel of the target subject (target voxel). In particular, we focused on the correspondence of the voxel coordinates with respect to the retinotopic map, between the source and target subjects. Voxels were sorted by the eccentricity and the polar angle coordinate of the retinotopic map, and the magnitudes of the weights of source voxels were averaged at each eccentricity or polar angle of source and target voxels. The weight distributions show relatively higher weight values along the diagonal line for both the eccentricity and the polar angle coordinate for both hemispheres (Fig. 5). The results indicate that the neural code converter predicts the activity of each target voxel mainly from the source voxels located in retinotopically corresponding regions. The relatively low weight magnitudes found at small eccentricities (Fig. 5A) and around the 90/270 deg polar angle (Fig. 5B) may be due to the crosstalk of informative voxels across hemispheres.

Next we compared the prediction performance of the neural code converter based on sparse regression, to those of more conventional methods that only use anatomical information, one-to-one voxel mapping by correlation, retinotopy, or orthogonal transformation of multi-voxel space (see Methods).
Fig. 6 shows the comparison result of prediction performance between the neural code converter and control methods. The voxel-wise correlations (Fisher's Z-transformed) averaged across all trials (8 trials/run/image × 8 runs × 5 images) and all source–target pairs (6 source–target pairs) were significantly higher than zero in all methods (a total of 1920 correlation values were evaluated; t-test, $p < 0.05$). Among all methods, the anatomy-based conversion that did not utilize functional information showed the

worst performance. The performances of one-to-one voxel mapping methods utilizing functional information including correlation-based conversion and retinotopy-based conversion were significantly worse than the performance of the neural code converter that utilized multi-voxel patterns for the prediction (t-test, $p < 0.05$). The procrustean conversion, a less flexible linear transformation than the neural code converter, showed significantly worse performance than the neural code converter (t-test, $p < 0.05$).

The results show that the neural code converter based on sparse regression outperformed the other methods, suggesting the advantage of the flexible regression model that exploits multi-voxel patterns. For further evaluation, only the neural code converter was applied as an inter-subject conversion algorithm.

Visual image reconstruction from predicted brain activity patterns

The second evaluation utilized predicted fMRI activity patterns for visual image reconstruction. We compared images reconstructed from the predicted and measured fMRI activity patterns. Fig. 7 shows examples of images reconstructed from S1's measured brain activity and those reconstructed from S1's predicted brain activity with S2 as the source subject. Results show that presented images could be reconstructed from the predicted brain activity with a comparable level of quality to those from the measured brain activity.

We also performed a quantitative evaluation of the accuracy of the neural code converter by calculating the pixel-wise correlation between visual images reconstructed from the measured and the predicted brain activity. The averaged pixel-wise correlation calculated from the reconstruction results is shown for each pair of source–target subjects in Fig. 8. The average correlation was calculated from 320 (= 5 figure images × (8×8 combinations of trials)) pairs of reconstructed images. Results showed that the pixel-wise correlation was significantly higher than zero in all source–target pairs (t-test, $p < 0.05$). The average of the pixel-wise correlations across subjects was 0.553 ± 0.008 (mean ± 95% confidence interval). These results suggest that predicted activity patterns carry enough information to achieve visual image reconstruction.

Decoding models trained with predicted brain activity patterns

The third evaluation was to test whether decoding models trained with predicted fMRI activity patterns alone could achieve high decoding accuracies using the measured data. We used the predicted brain activity patterns to build a new decoder to classify brain activity into five categories corresponding to each of the figure images (eight trials for each image). The results showed high decoding accuracies significantly exceeding the chance level in all source–target pairs (binominal test, $p < 0.05$; average

across pairs, 78.333 ± 18.484%; Fig. 9 A). The classification performance was comparable to that of the decoder trained with measured brain activity (94.167 ± 5.204%; Fig. 9B). The results suggest that once a neural code converter is established, decoding models can be trained using the predicted fMRI activity patterns, without the need for actual measurements.

Discussion

In this paper, we proposed an inter-subject neural code converter that can translate brain activity patterns of one subject to another subject's, while preserving representational contents. The converter trained with a limited number of random images successfully predicted fMRI activity patterns for unseen visual images (Fig. 4). We further demonstrated that visual images could be reconstructed from predicted fMRI activity patterns (Fig. 7 and Fig. 8). The predicted fMRI activity patterns could be also used to train accurate classifiers for measured fMRI activity patterns (Fig. 9). These results suggest that our neural code converter may be able to predict brain activity patterns corresponding to arbitrary perceptual states by only conducting experiments to collect data necessary for training the neural code converter.

An important feature of the neural code conversion that provides advantages over previous methods is the ability to exploit multi-voxel patterns in fMRI activity in a flexible way. Although structural differences in brain anatomy can be standardized by anatomical normalization, individual differences in fMRI activity patterns were not able to be assimilated accurately for voxels with the same anatomical coordinate (Fig. 6). This suggests that inter-subject differences in fMRI activity patterns cannot be explained by anatomical differences alone. When spatial positions of voxels were replaced so that fMRI activity patterns of paired voxels between subjects were most correlated, fMRI activity patterns were able to be partially assimilated between the paired subjects, showing a significant positive correlation. However, the accuracy achieved by the voxel replacement was still significantly lower than that achieved by the neural code conversion (Fig. 6). Thus, inter-subject differences in fMRI activity patterns are not simply explainable as a result of individual variations in the spatial arrangement of cortical regions that have similar response characteristics. Although Fig. 5 shows that the neural code converter tends to use retinotopically corresponding voxels, it also uses voxels with a bit broader range of coordinates. It is also known that accurate prediction of the contrast at a single visual field location requires a complex response pattern of voxels with a range of retinotopic coordinates (Miyawaki et al., 2008). Together, one-to-one mapping based on the positional correspondence seems to be insufficient to achieve accurate prediction of one subject's fMRI activity patterns from another's.

The accuracy of the neural code conversion was also demonstrated by visual image reconstruction that utilized predicted fMRI activity patterns. Pixel-wise correlations between images reconstructed from measured and predicted fMRI activity patterns were calculated (Fig. 8). Pixel-wise correlations (0.553 ± 0.008 (mean ± 95% confidence interval)) were higher than voxels-wise correlations (0.399 ± 0.014 (mean ± 95% confidence interval)). This may be because the reconstruction model assigns weights of greater magnitudes to voxels that are more selective to stimuli (Miyawaki et al., 2008). Pixel patterns calculated

from selectively weighted voxels may represent more reliable responses to stimuli than non-weighted voxel patterns.

Machine learning-based brain decoding often requires acquisition of a large number of data sets to train a model customized to each subject. Our method may partially resolve this issue by utilizing predicted fMRI activity patterns to train decoders for a person who does not undergo actual experiments. The neural code converter could predict fMRI activity patterns of the target subject for unseen visual stimuli given the fMRI activity patterns of another subject, and the predicted fMRI activity patterns can be used to train a decoder to classify fMRI activity patterns measured for a target subject (Fig. 9). This suggests that once a neural code converter is trained, fMRI activity patterns necessary to create decoders for a target subject can be substituted by those measured from another subject. Thus, experimental data necessary for a target subject are only for training a neural code converter, which can be used to predict activity patterns for other subjects, without performing actual experiments. If this method is applied to brain–machine interfaces based on neural decoding, it will potentially reduce the user's load required for collecting data for system caliblration.

Although we tested the neural code converter on fMRI data from the early visual cortex in this study, the framework could also be applied to the higher visual cortices and even other sensory areas, such as the auditory cortex and the somatosensory cortex. A previous study (Hasson et al., 2004) has shown that brain activity patterns from different subjects were significantly correlated across broad areas of the brain, including visual areas, auditory areas, and areas around the superior temporal sulcus and the lateral sulcus), while subjects viewed identical visual stimuli (e.g., movie stimuli). Thus, our method may also be applicable to establish a systematic mapping of brain activity patterns between subjects for these areas. Depending on the brain area of interest, other machine learning methods, which are based on hypotheses about information representation on neural population, might be more appropriate. For example, if information is represented over multiple voxels in a broadly distributed manner, a non-sparse method might show better performance.

An interesting application would be to use the neural code converter to analyze individual differences in brain activity patterns relevant to human complex functions such as higher cognitions, social behaviors, preferences, personalities, and cultural backgrounds. The neural code converter provides a mapping of brain activity patterns from one person to another, so the mapping parameters (e.g., weight values) could give an insight into how differently the brain represents information relevant for such complex functions between different subjects.

A more advanced application of the neural code converter could be enabling direct brain-to-brain communication. If the converted brain activity is used for designing a brain stimulation pattern, the stimulation might induce a similar mental content. In the last decade, new technologies as such as optogenetics (Deisseroth, et al., 2006) have dramatically improved the capacity of brain stimulation in both spatial and temporal resolution. Furthermore, a proof-of-concept of brain-to-brain communication has been demonstrated by using EEG and TMS (Grau et al., 2014), although the system only works at an anatomically-coarse scale. Since our neural code converter can deal with information represented in fine-grained activity patterns, high-resolution brain stimulation designed with converted patterns may allow for the transmission of detailed mental contents between two persons. This could create a new dynamic in how humans communicate with each other.

Acknowledgements

The authors thank T. Horikawa for technical assistance, M. Takemiya, P. Sukhanov and K. Majima for editing the manuscript.

References

Bishop C., 2006. Pattern Recognition and Machine Learning. Springer.

Chadwick M., Bonnici H., Maguire E., 2014. CA3 size predicts the precision of memory recall. Proc Natl Acad Sci U S A. 111, 10720-10725.

Deisseroth K., Feng G., Majewska A., Miesenbo¨ck G., Ting A., and Schnitzer M., 2006. Next-Generation Optical Technologies for Illuminating Genetically Targeted Brain Circuits. The Journal of Neuroscience, October 11, 26, 10380 –10386

Engel S.A., Rumelhart D.E., Wandell B.A., Lee A.T., Glover G.H., Chichilnisky E., and Shadlen M.N., 1994. fMRI of human visual cortex. Nature 369, 525.

Evans A.C., Marrett S., Neelin P., Collins L., Worsley K., Dai W., Milot S., Meyer E., Bub D., 1992. Anatomical mapping of functional activation in stereotactic coordinate space. Neuroimage 1, 43-53.

Evans A.C., Collins D.L., Mills S.R., Brown E.D., Kelly R.L., Peters T.M., 1993. 3D statistical neuroanatomical models from 305 MRI volumes. Nuclear Science Symposium and Medical Imaging Conference, 1993: 1993 IEEE Conference Record, SanFrancisco, CA, USA, 1813–1817.

Fischl B., Rajendran N., Busa E., Augustinack J., Hinds O., Yeo B.T., Mohlberg H., Amunts K., Zilles K., 2008. Cortical folding patterns and predicting cytoarchitecture. Cereb. Cortex18, 1973-1980.

Friston K, Chu C, Mourão-Miranda J, Hulme O, Rees G, Penny W, Ashburner J., 2008. Bayesian decoding of brain images. Neuroimage. 39, 181-205.

Grau C., Ginhoux R., Riera A., Nguyen T., Chauvat H., Berg M., Amengual J., Pascual-Leone A, Ruffini G., 2014. Conscious Brain-to-Brain Communication in Humans Using Non-Invasive Technologies. PLOS ONE, August 19.

Hasson U., Nir Y., Levy I., Fuhrmann G., and Malach R., 2004. Intersubject synchronization of cortical activity during natural vision. Science 303, 1634-1640.

Haxby J., Guntupalli J., Connolly A., Halchenko Y., Conroy B., Gobbini M.,x Hanke M., and Ramadge P., 2011 A Common, High-Dimensional Model of the Representational Space in Human Ventral Temporal Cortex. Neuron 72, 404–416.

Kamitani Y., and Tong F., 2005. Decoding the visual sand subjective contents of the human brain. Nature Neuroscience 8, 679-685.

Miyawaki Y., Uchida H., Yamashita O., Sato M., Morito Y., Tanabe H., Sadato N., and, Kamitani Y., 2008. Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. Neuron 60, 915-929.

Sabuncu M., Singer B., Conroy B.,x Bryan B., Ramadge P., and Haxby J., 2009. Function based inter-subject alignment of human cortical anatomy. Cerebral Cortex 20, 130-140.

Sato, M.A., 2001. Online model selection based on the variational Bayes. Neural Comp. 13, 1649-1681.

Sato, M.A., Yoshioka, T., Kajihara, S., Toyama, K., Goda, N., Doya, K., Kawato, M., 2004. Hierarchical Bayesian estimation for MEG inverse problem. NeuroImage 23, 806-826.

Sereno M., Dale A., Reppas J., Kwong K., Belliveau J., Brady T., Rosen B., and Tootell R., 1995. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science 268, 889-893.

Talairach J., and Tournoux P., 1988. Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system—an approach to cerebral imaging. Thieme Medical Publishers..

Ting, J.A., D'Souza, A., Yamamoto, K., Yoshioka, T., Hoffman, D.S., Kakei, S., Lauren, S., Kalaska, J.F., Kawato, M., Strick, P.L., Schaal, S., 2005. Predicting EMG data from M1 neurons with Variational Bayesian least squares. Proceedings of Advances in Neural Information Processing Systems 18 (NIPS2005). MIT press, Cambridge, MA.

Ting, J.A., D'Souza, A., Yamamoto, K., Yoshioka, T., Hoffman, D.S., Kakei, S., Lauren, S., Kalaska, J.F., Kawato, M., Strick, P.L., Schaal, S., 2008. Variational Bayesian least squares: an application to brain machine interface data. Neural Networks 21, 1112-1131.

Toda A., Imamizu H., Kawato M., and Sato M., 2011. Reconstruction of two-dimensional movement trajectories from selected magnetoencephalography cortical currents by combined sparse Bayesian methods. NeuroImage 54, 892-905.

Van Essen D.C., 2004. Surface-based approaches to spatial localization and registration in primate cerebral cortex. Neuroimage 23, S97-S107.

Van Essen D.C., 2005. A population-average, landmark- and surface-based (PALS) atlas of human cerebral cortex. Neuroimage 28, 635-662.

Yamada K., Miyawaki Y., and Kamitani Y., 2011. Neural Code Converter for Visual Image Representation. IEEE International Workshop on Pattern Recognition in NeuroImaging, 37-40.

Yamashita O., Sato M., Yoshioka T., Tong F., Kamitani Y., 2008. Sparse estimation automatically selects voxels relevant for the decoding of fMRI activity patterns. Neuroimage 42, 1414–1429.

Figure legends

Fig. 1. Schematics of the neural code converter. (A) Training the neural code converter. An identical stimulus set, consisting of random-patterned images, is presented to a pair of subjects. Then, the pairs of fMRI activity patterns are used to train linear regression models to predict fMRI signals for each voxel of the target subject, given the fMRI signal patterns of the source subject. The set of the linear regression models to predict fMRI signals for all voxels of the target subject constitutes the neural code converter. (B) Prediction of fMRI activity patterns through neural code conversion. fMRI activity patterns from the source subject evoked by visual stimuli independent of those used for the training are converted to fMRI activity patterns for the target subject.

Fig. 2. Visual image reconstruction through neural code conversion. Visual stimulus images are reconstructed from the target subject's predicted fMRI activity patterns.

Fig. 3. Training a decoder using fMRI activity patterns predicted through neural code conversion. A decoder (classification model to predict presented image labels) for the target subject is trained with fMRI activity patterns for the target subject, predicted by the neural code converter. fMRI activity patterns of the target subject, measured for the same visual image set, are then classified into one of five classes by the decoder trained by the neural code converter.

Fig. 4. Voxel-wise correlations between measured and predicted fMRI activity patterns through neural code conversion. Correlation values were converted into Fisher's z-scores. Six source–target pairs are shown (mean ± 95% confidence interval). Dotted lines indicate keupper bounds of voxel-wise correlations (averaged over stimulus pairs).

Fig. 5. Voxel weight distributions from the source subject to predict fMRI activity patterns for the target subject. Voxels for the target subjects are sorted by (A) the eccentricity, and (B) the polar angle coordinate of the retinotopic map (horizontal axis, 0.67 deg bins for eccentricity and 15 deg bins for polar angle) of each hemisphere. Weight values of the source subject are also sorted by the coordinates of the retinotopic map for the corresponding voxels (vertical axis, 0.67 deg bins for eccentricity and 15 deg bins for polar angle) for each hemisphere. The magnitude of voxel weights was averaged in each target voxel location and source voxel location (six source–target pairs from three subjects pooled). White cells denote no corresponding voxel with non-zero weight.

Fig. 6. Performance comparison (voxel-wise correlation) between the neural code converter and

conventional methods (conversion by anatomical normalization, voxel position replacement based on temporal correlation, correspondence of retinotopic maps, and procrustean multi-voxel transformation). Correlation values were converted into Fisher's z-scores and then averaged across all trials for all source–target pairs (mean ± 95% confidence interval). Dotted lines indicate upper bounds of voxel-wise correlations (averaged over subjects and stimulus pairs).

Fig. 7. Visual stimulus images (contrast patterns) reconstructed from measured and predicted fMRI activity patterns of S1 (S2 as source subject). Results presented here are examples of a single run of the figure image session consisting of ten trials of the stimulus presentation (presentation order (left to right) preserved).

Fig. 8. Pixel-wise correlations between images reconstructed from fMRI activity patterns using measured and predicted (through neural code conversion) fMRI activity. Correlation values were converted into Fisher's z-scores. Six source–target pairs are shown (mean ± 95% confidence interval). Dotted lines indicate upper bounds of pixel-wise correlations (averaged over stimulus pairs).

Fig. 9. Classification performance of decoders trained with (A) predicted, and (B) measured fMRI activity patterns.

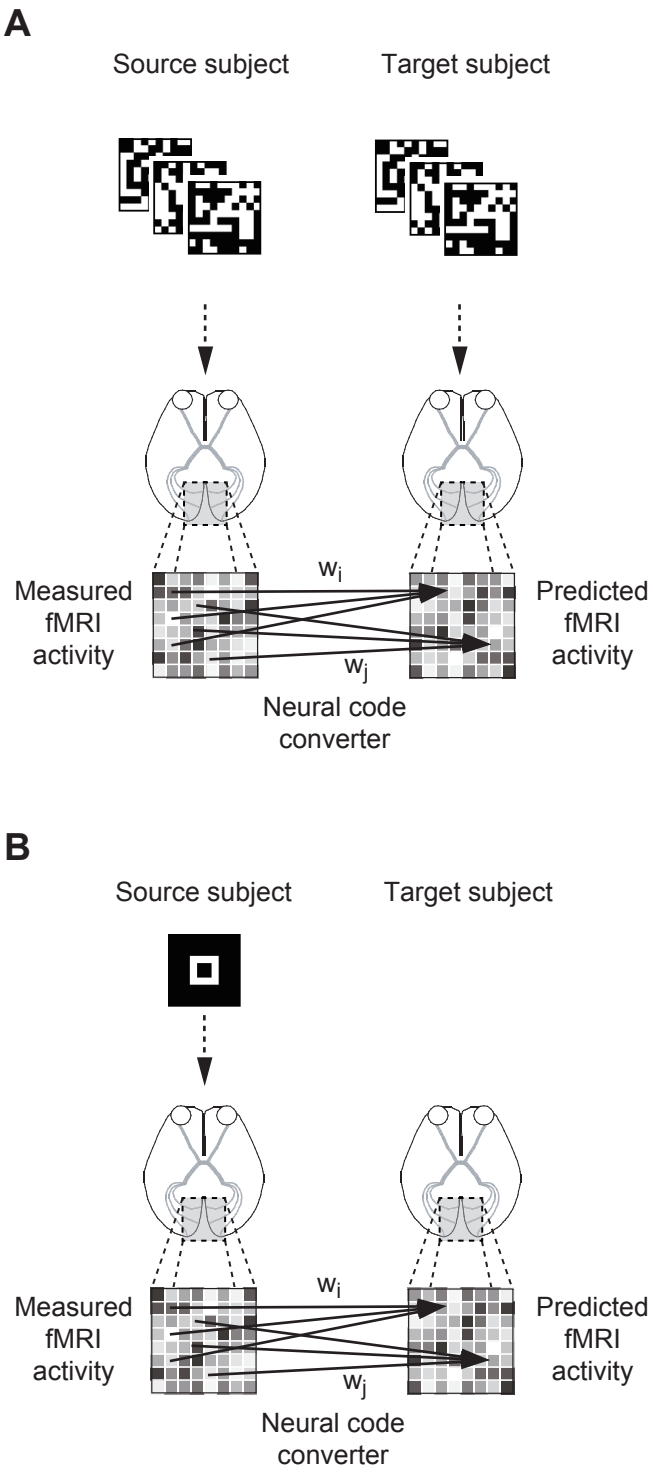

Fig. 1. Schematics of the neural code converter. (A) Training the neural code converter. An identical stimulus set, consisting of random-patterned images, is presented to a pair of subjects. Then, the pairs of fMRI activity patterns are used to train linear regression models to predict fMRI signals for each voxel of the target subject, given the fMRI signal patterns of the source subject. The set of the linear regression models to predict fMRI signals for all voxels of the target subject constitutes the neural code converter. (B) Prediction of fMRI activity patterns through neural code conversion. fMRI activity patterns from the source subject evoked by visual stimuli independent of those used for the training are converted to fMRI activity patterns for the target subject.

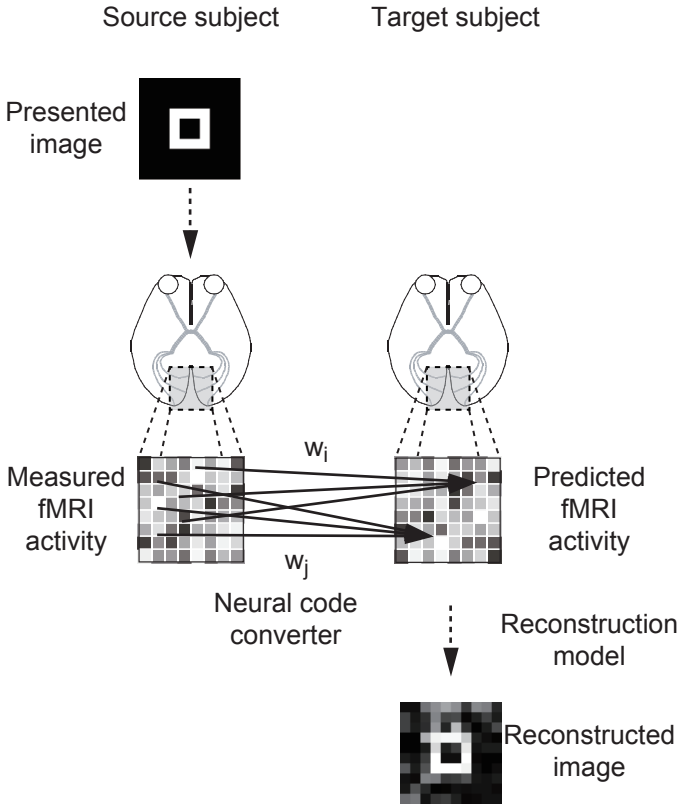

Fig. 2. Visual image reconstruction through neural code conversion. Visual stimulus images are reconstructed from the target subject's predicted fMRI activity patterns.

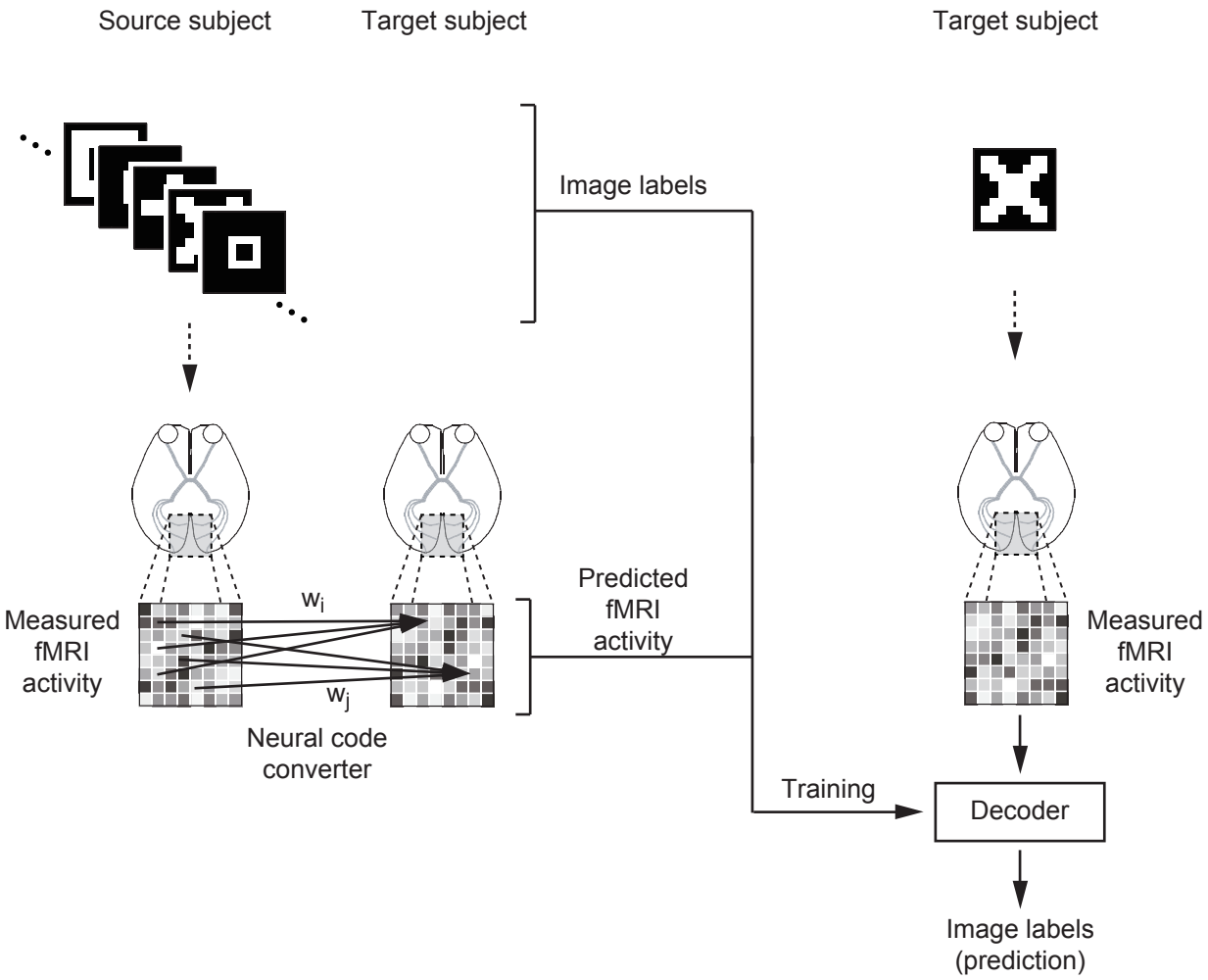

Fig. 3. Training a decoder using fMRI activity patterns predicted through neural code conversion. A decoder (classification model to predict presented image labels) for the target subject is trained with fMRI activity patterns for the target subject, predicted by the neural code converter. fMRI activity patterns of the target subject, measured for the same visual image set, are then classified into one of five classes by the decoder trained by the neural code converter.

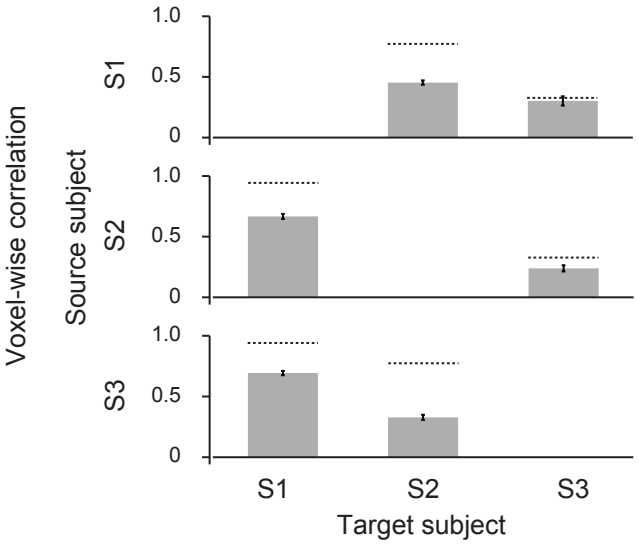

Fig. 4. Voxel-wise correlations between measured and predicted fMRI activity patterns through neural code conversion. Correlation values were converted into Fisher's z-scores. Six source–target pairs are shown (mean ± 95% confidence interval). Dotted lines indicate upper bounds of voxel-wise correlations (averaged over stimulus pairs).

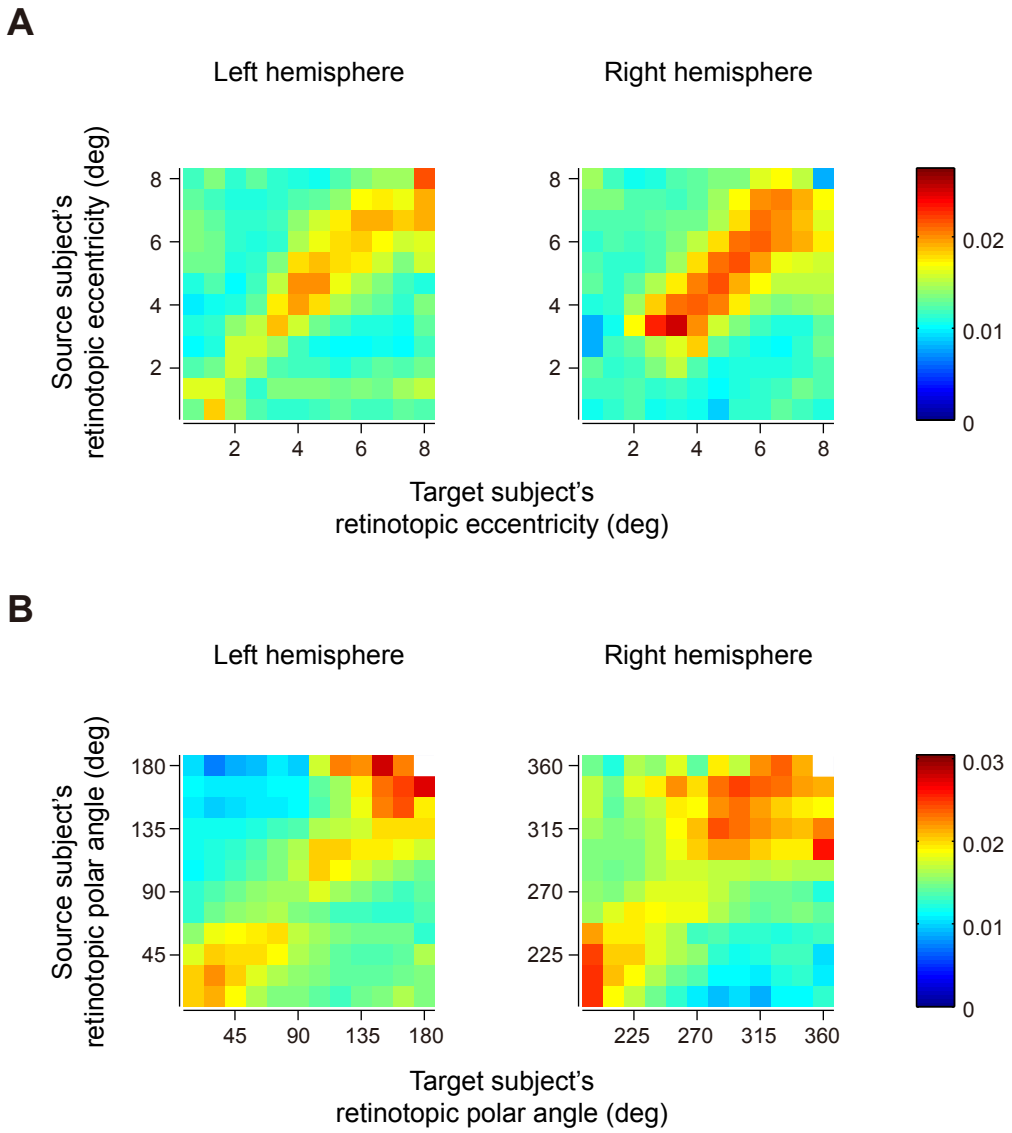

Fig. 5. Voxel weight distributions from the source subject to predict fMRI activity patterns for the target subject. Voxels for the target subjects are sorted by (A) the eccentricity, and (B) the polar angle coordinate of the retinotopic map (horizontal axis, 0.67 deg bins for eccentricity and 15 deg bins for polar angle) of each hemisphere. Weight values of the source subject are also sorted by the coordinates of the retinotopic map for the corresponding voxels (vertical axis, 0.67 deg bins for eccentricity and 15 deg bins for polar angle) for each hemisphere. The magnitude of voxel weights was averaged in each target voxel location and source voxel location (six source–target pairs from three subjects pooled). White cells denote no corresponding voxel with non-zero weight.

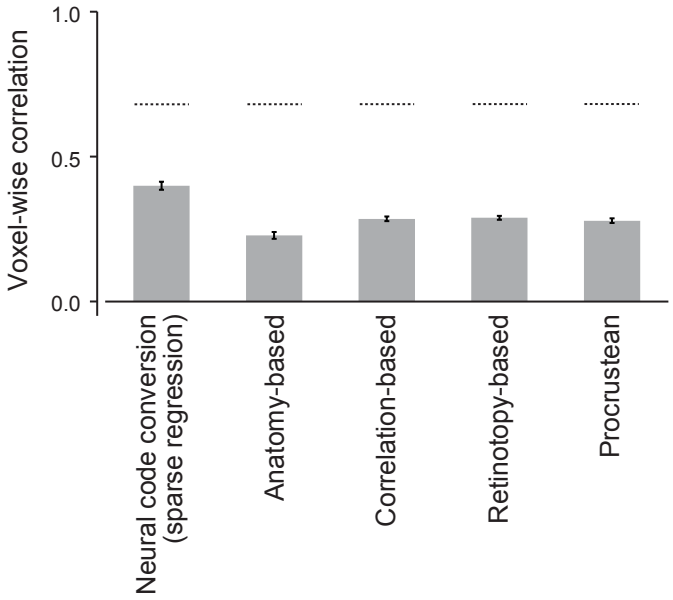

Fig. 6. Performance comparison (voxel-wise correlation) between the neural code converter and conventional methods (conversion by anatomical normalization, voxel position replacement based on temporal correlation, correspondence of retinotopic maps, and procrustean multi-voxel transformation). Correlation values were converted into Fisher's z-scores and then averaged across all trials for all source–target pairs (mean ± 95% confidence interval). Dotted lines indicate upper bounds of voxel-wise correlations (averaged over subjects and stimulus pairs).

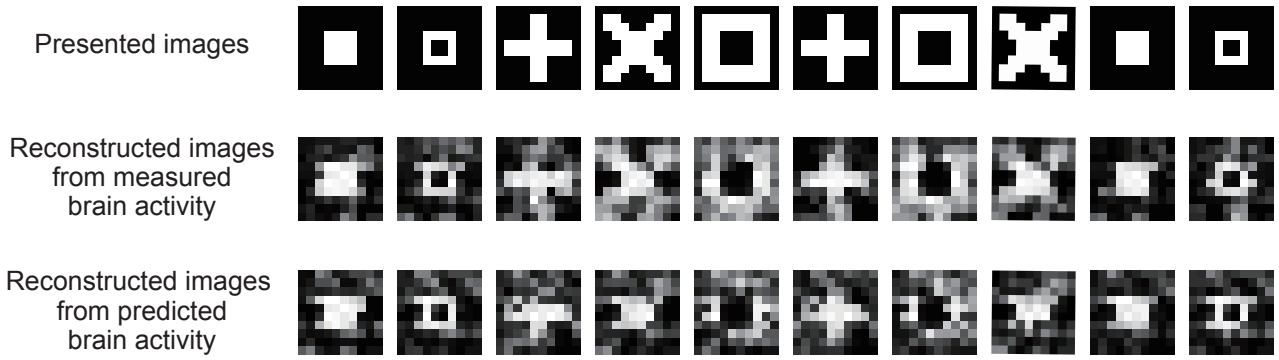

Fig. 7. Visual stimulus images (contrast patterns) reconstructed from measured and predicted fMRI activity patterns of S1 (S2 as source subject). Results presented here are examples of a single run of the figure image session consisting of ten trials of the stimulus presentation (presentation order (left to right) preserved).

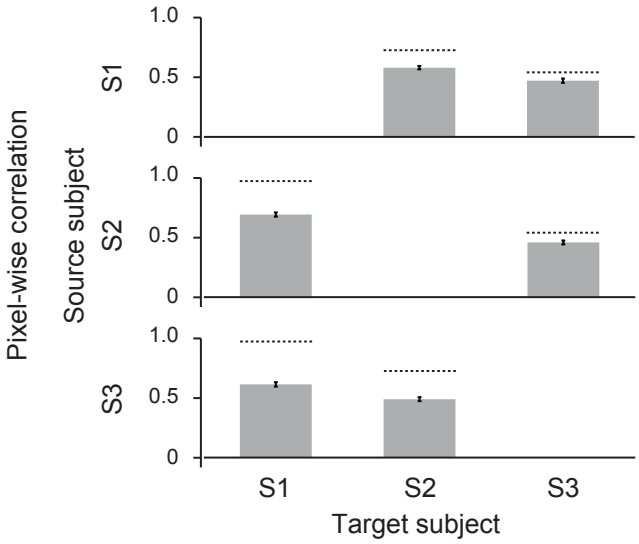

Fig. 8. Pixel-wise correlations between images reconstructed from fMRI activity patterns using measured and predicted (through neural code conversion) fMRI activity. Correlation values were converted into Fisher's z-scores. Six source–target pairs are shown (mean ± 95% confidence interval). Dotted lines indicate upper bounds of pixel-wise correlations (averaged over stimulus pairs).

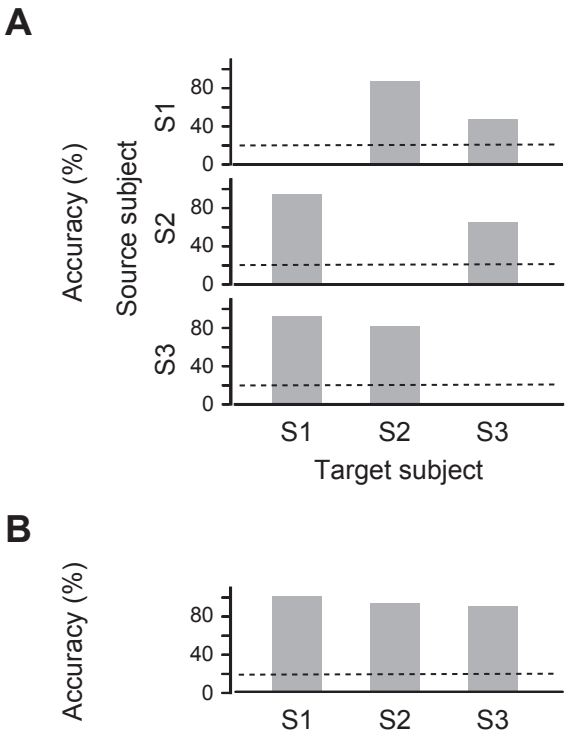

Fig. 9. Classification performance of decoders trained with (A) predicted, and (B) measured fMRI activity patterns.